Title
Hierarchical complexity and the size limits of life

Journal
Proceedings of the Royal Society B, 284(1857)

ISSN
0962-8452

Authors
Heim, Noel A
Payne, Jonathan L
Finnegan, Seth
et al.

Publication Date
2017-06-28

DOI
10.1098/rspb.2017.1039

Peer reviewed

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Heim NA *et al*. 2017 Hierarchical complexity and the size limits of life. *Proc. R. Soc. B* **284**: 20171039. http://dx.doi.org/10.1098/rspb.2017.1039

Received: 12 May 2017
Accepted: 25 May 2017

Subject Category:
Evolution

Subject Areas:
evolution, palaeontology

Keywords:
body size, evolution, hierarchy, complexity, macroecology, macroevolution

Author for correspondence:
Noel A. Heim
e-mail: naheim@stanford.edu

Electronic supplementary material is available online at http://dx.doi.org/10.6084/m9.figshare.c.3799456.

THE ROYAL SOCIETY PUBLISHING

Hierarchical complexity and the size limits of life

Noel A. Heim[1], Jonathan L. Payne[1], Seth Finnegan[2], Matthew L. Knope[3], Michał Kowalewski[4], S. Kathleen Lyons[5], Daniel W. McShea[6], Philip M. Novack-Gottshall[7], Felisa A. Smith[8] and Steve C. Wang[9]

[1]Department of Geological Sciences, Stanford University, Stanford, CA 94305, USA
[2]Department of Integrative Biology, University of California, Berkeley, CA 94720, USA
[3]Department of Biology, University of Hawaii, Hilo, HI 96720, USA
[4]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
[5]School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE 68588, USA
[6]Department of Biology, Duke University, Durham, NC 27708, USA
[7]Department of Biological Sciences, Benedictine University, Lisle, IL 60532, USA
[8]Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA
[9]Department of Mathematics and Statistics, Swarthmore College, Swarthmore, PA 19081, USA

NAH, 0000-0002-4528-345X; JLP, 0000-0002-9601-3310; DWM, 0000-0001-9398-0025; SCW, 0000-0002-8953-285X

Over the past 3.8 billion years, the maximum size of life has increased by approximately 18 orders of magnitude. Much of this increase is associated with two major evolutionary innovations: the evolution of eukaryotes from prokaryotic cells approximately 1.9 billion years ago (Ga), and multicellular life diversifying from unicellular ancestors approximately 0.6 Ga. However, the quantitative relationship between organismal size and structural complexity remains poorly documented. We assessed this relationship using a comprehensive dataset that includes organismal size and level of biological complexity for 11 172 extant genera. We find that the distributions of sizes within complexity levels are unimodal, whereas the aggregate distribution is multimodal. Moreover, both the mean size and the range of size occupied increases with each additional level of complexity. Increases in size range are non-symmetric: the maximum organismal size increases more than the minimum. The majority of the observed increase in organismal size over the history of life on the Earth is accounted for by two discrete jumps in complexity rather than evolutionary trends within levels of complexity. Our results provide quantitative support for an evolutionary expansion away from a minimal size constraint and suggest a fundamental rescaling of the constraints on minimal and maximal size as biological complexity increases.

1. Introduction

The size of the largest organism on the Earth has increased by approximately 18 orders of magnitude over the course of the Geozoic (approx. 3.8 billion years ago (Ga)–present [1]). Much of this increase occurred during two major jumps associated with the origin of eukaryotic cells approximately 1.9 Ga and animals approximately 0.6 Ga [2]. Though increases in the largest known organisms have been well documented [2], changes in the overall distribution of organismal sizes over the Geozoic remain poorly characterized. The evolutionary increases in size correspond to increases in biological complexity. There are two forms of complexity. The first is defined by structural hierarchy [3], or 'vertical' complexity [4], which is the number of levels of nestedness or levels of organization in an organism. For example, solitary eukaryotic cells arose historically as an association of prokaryotic cells [5] and are therefore one vertical level above prokaryotes. A multicellular eukaryotic organism is an association of unicellular protists [6] and is therefore one level higher yet. Vertical complexity contrasts with 'horizontal' complexity, which is the number of part types within a given level, such as the number of

cell types within an animal [7]. Here, we use the word 'complexity' in its vertical, anatomical sense.

The qualitative trajectory of maximum organismal size over the history of life suggests a connection between complexity and size. However, in the absence of data describing the full distribution of organismal sizes, the precise nature of the relationship between complexity and other aspects of the size distribution, such as the minimum, median, mean and range, has remained largely unexplored. Furthermore, there remains the question of whether the observed increases in maximum size [2] and complexity [3,8] are the results of driven or passive evolutionary processes [9].

Stanley [10] (see also Gould [11,12]) argued that trends such as increases in mean and maximum organismal size were best explained as increases in variance in a system dominated by passive processes and bounded by a minimum size, rather than driven trends reflecting selective advantages associated with larger size. This increasing-variance hypothesis predicts a diffusion-like evolutionary process with a single constraint on absolute minimum size. If this process operates, organismal size should appear to diffuse away from a single absolute minimum size, and the diffusion should be independent of complexity level. A further prediction is that all of life, in aggregate, follows a right-skewed, unimodal size–frequency distribution [12]. Alternatively, Knoll & Bambach [13] described the history of life as a sequence of evolutionary megatrajectories, or a linked series of discrete trends, that have played out during the Geozoic. Their megatrajectories are not purely structural—as vertical complexity is—but also have an ecological dimension.

Adapting their hypothesis to the vertical complexity case, the expectation is that the variance in size within each megatrajectory increased over geologic time, but is bounded by both minimum and maximum constraints imposed by hierarchical structure at each level. For example, the smallest possible prokaryotic cell—a size minimum for that level—is thought to require a diameter of at least 250 nm, just large enough for a minimum number of ribosomes to mediate gene expression [14]. At the other extreme, a prokaryotic cell cannot become so large that metabolites and essential biomolecules cannot reach the cell's interior through diffusion [15], suggesting that increased vertical complexity was a prerequisite for evolving organisms substantially larger than a prokaryote. The megatrajectory hypothesis [13] further postulates that episodic increases in maximum complexity occurred over time as constraints on the maximum were occasionally and successively surmounted by organisms possessing key evolutionary innovations.

This suggestion is different from, and independent of, earlier findings and theorizing on complexity in the horizontal sense, number of part types. For example, earlier studies examining the relationship between size and horizontal complexity found a log-linear relationship between size and the number of cell types in clades as diverse as animals, plants and multicellular algae [16]. Bonner [17] suggested that this log-linearity represented a trade-off between size and internal physiological efficiency.

Here we quantify the relationship between vertical complexity and the distribution of organismal sizes across a wide range of organisms, from viruses to multicellular eukaryotes. Using a dataset of organismal sizes categorized into four levels of complexity, we compare two models that require increased complexity as a prerequisite for evolving organisms substantially larger than a prokaryote. In the first model, based on Stanley [10] and Gould [11,12], the evolution of organismal size diffuses away from a single absolute minimum size, is independent of complexity level, and is characterized by a unimodal, right-skewed size distribution. In the second model, based on Knoll & Bambach [13], size evolution is dependent upon the level of vertical complexity, and each complexity level is characterized by a unimodal size distribution bound by both a minimum and maximum size.

2. Data and methods

We classify organisms into four levels of vertical complexity: (i) virus, (ii) single-celled prokaryote, (iii) single-celled eukaryote and (iv) non-colonial multicellular eukaryote. The main advantage of our framework is that it is purely structural and can, in principle, be expanded and applied across all levels of biological organization from organic molecules to planetary biospheres [3]. Because the criteria for each level are strictly structural, assignment of a taxon to a level of complexity can be done objectively and unambiguously for the vast majority of organisms. We avoid using mixed criteria that confound structure, function and ecology. The hierarchy used here allows direct comparison among levels and avoids concepts that might seem to be complexity related but are difficult to quantify, such as intelligence.

We assessed the relationship between vertical complexity and organismal size using a dataset that includes biovolume (in units of cubic millimetres) for 11 172 living genera. Our data span all four levels of complexity and include all three domains of life plus viruses. Sizes were compiled at the genus level, using the holotype to estimate biovolume when possible. We chose genera as our operational units to facilitate data acquisition and to minimize the statistical noise associated with identifying species [18]. Size distributions can be based on body size estimates compiled using various operational units (specimens, species types, genus types, etc.). However, at the broadest scale of domains, the choice of the operational unit is inconsequential because subtle differences in taxonomic resolution at the finest scales, as well as inconsistencies in how those units are defined across domains, are analytically negligible.

Viruses, which are included here, are not named and organized within the Linnaean classification system; thus, taxonomy did not guide our data collection for viruses. Rather, we compiled virus capsid sizes without regard for formal classification. Nevertheless, our virus compilation includes single- and double-stranded forms of DNA and RNA viruses. We included mainly viruses that infect plants, and also some that infect mammals and invertebrates.

Biovolume was estimated as a three-dimensional ellipsoid based on linear measurements of the three primary body/cell/capsid axes taken from illustrated specimens or listed in the text of publications. To ensure we captured the total organismal size range, we included the largest and smallest known organisms for each vertical level (see the electronic supplementary material for size sources and data). We examined the distribution of sizes within each complexity level, noting the centre and range of size within each distribution. Biovolume was $\log_{10}$ transformed prior to all analyses. A Brown–Forsythe test of equal variances was used to test for significant differences in variance among vertical levels

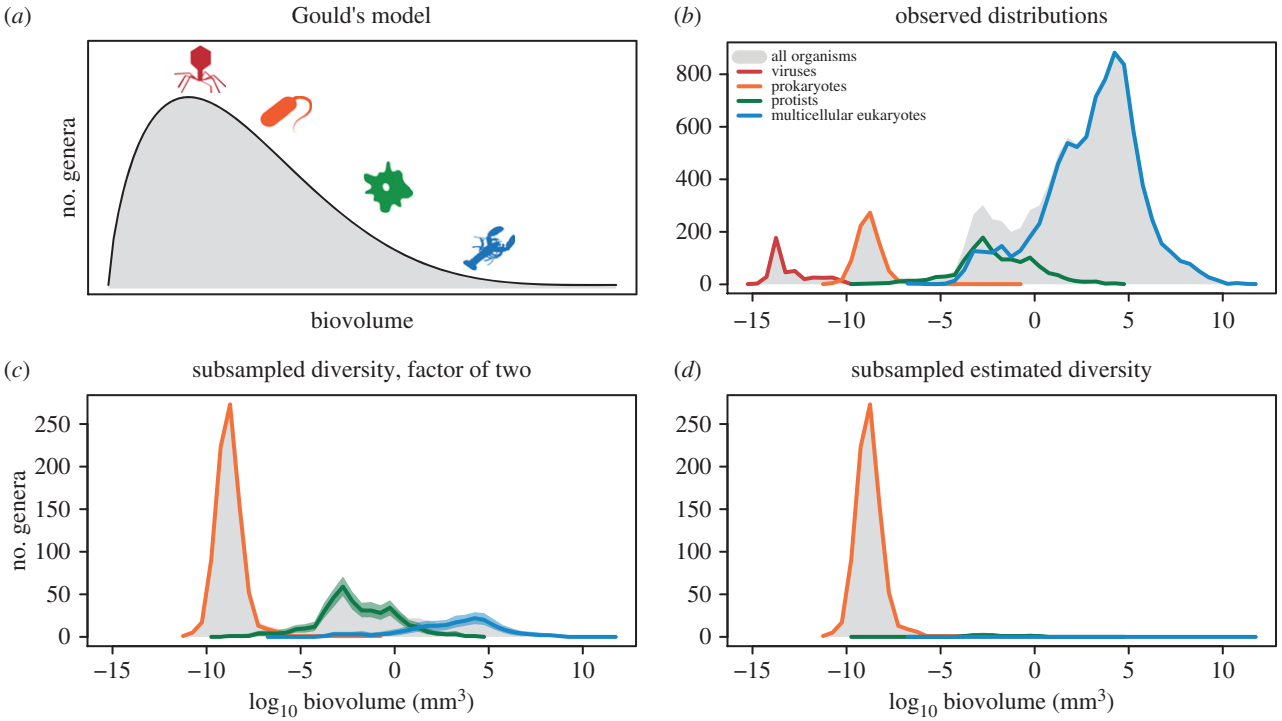

Figure 1. Aggregated distributions of organismal size. (*a*) The hypothetical unimodal right-skewed distribution of organismal sizes expected under the Gould [12] model. (*b*) The observed distribution of organismal sizes in our data. The grey area highlights the cumulative distribution of sizes and the coloured lines correspond to the distributions of the individual vertical levels of complexity. (*c*) Subsampled distributions where protists and multicellular eukaryotes are assumed to be one-half and one-quarter as diverse, respectively, as prokaryotes. (*d*) Subsampled distributions where the relative diversities are assumed to be proportional to the estimates for prokaryotes, protists and multicellular eukaryotes [18]. Subsampling our data does not produce the unimodal distribution predicted by Gould [12]. Grey areas of (*c*) and (*d*) are the 50th percentile of 10 000 subsampled distributions. The solid coloured lines are the 50th percentiles of the individual levels and the coloured areas bound the 5th and 95th percentiles. Regardless of subsampling, the overall distribution remains multimodal.

using the *lawstat* [19] package for R [20]. We tested for multimodality with Hartigan's dip test [21] using the *diptest* package [22] for R [20].

We tested for sampling effects on the modality of our data by subsampling each complexity level such that true diversities (i.e. species richness) were proportionally represented; we assume that genus richness is proportional to species richness [23]. Our largest sample is for multicellular eukaryotes followed by protists, prokaryotes and viruses. We note that (i) the proportion of extant species that have been discovered and described is relatively low, (ii) extrapolating the planet's true biodiversity is not straightforward and (iii) these estimates are imprecise [24,25]. These difficulties, combined with the fact that diversity estimates of viruses, prokaryotes and eukaryotes are typically conducted separately and with different methodologies, makes determining the precise relative diversities of these groups difficult. We used subsampling with replacement to compare among-level size distributions in a way that more realistically reflects the true relative diversities of the groups sampled. Owing to the small sample of virus sizes in our data (405 viral 'species') and their large estimated diversity of approximately 100 000 000 genomic species [26], they have been excluded from subsampling. Including the viruses as the most diverse group in subsamples would force subsamples of the other levels to be too small to meaningfully interpret. Subsampling was done with two different treatments of the relative diversity of prokaryotes, solitary eukaryotes and multicellular eukaryotes. For the first (factor-of-two) subsampling treatment, we assumed that prokaryotes are the most diverse group and that unicellular eukaryotes were half as diverse as the prokaryotes and that multicellular eukaryotes are half as diverse as unicellular eukaryotes. This simple model is consistent with Gould's [12] qualitative perception of diversity. The use of other scaling factors produces similar results to those reported here. The second (estimated diversity) subsampling treatment assumed that there are 3.83 million species of prokaryotes, 1.5 million species of protists, and 56 000 species of multicellular eukaryotes [27]. The prokaryotes have the smallest sample size ($n = 851$) in our dataset; thus, their observed distribution was used without subsampling. Subsampling to these relative diversities was carried out 10 000 times.

3. Results

The size–frequency distribution of each vertical level is approximately unimodal. However, there is a hint of bimodality in the protists (figure 2*b*; electronic supplementary material, figure S1) that reflects a slight over-representation of Foraminifera, one of the most diverse groups of protists, in our dataset. However, the degree of bimodality is low and the distributions of Foraminifera and non-foraminiferan protists largely overlap (electronic supplementary material, figure S4).

The aggregate size distribution of all organisms is strongly multimodal (figure 1*b*); Hartigan's dip test rejects the null hypothesis of unimodality ($D = 0.02$, $p \ll 0.001$). This multimodality persists so long as protists and animals constitute a non-negligible fraction of total diversity. The first treatment (factor of two), where diversity in higher complexity groupings is assumed to be half that of lower levels, remains significantly multimodal (figure 1*c*). The second treatment (estimated diversity), where distributions are adjusted to match estimates of

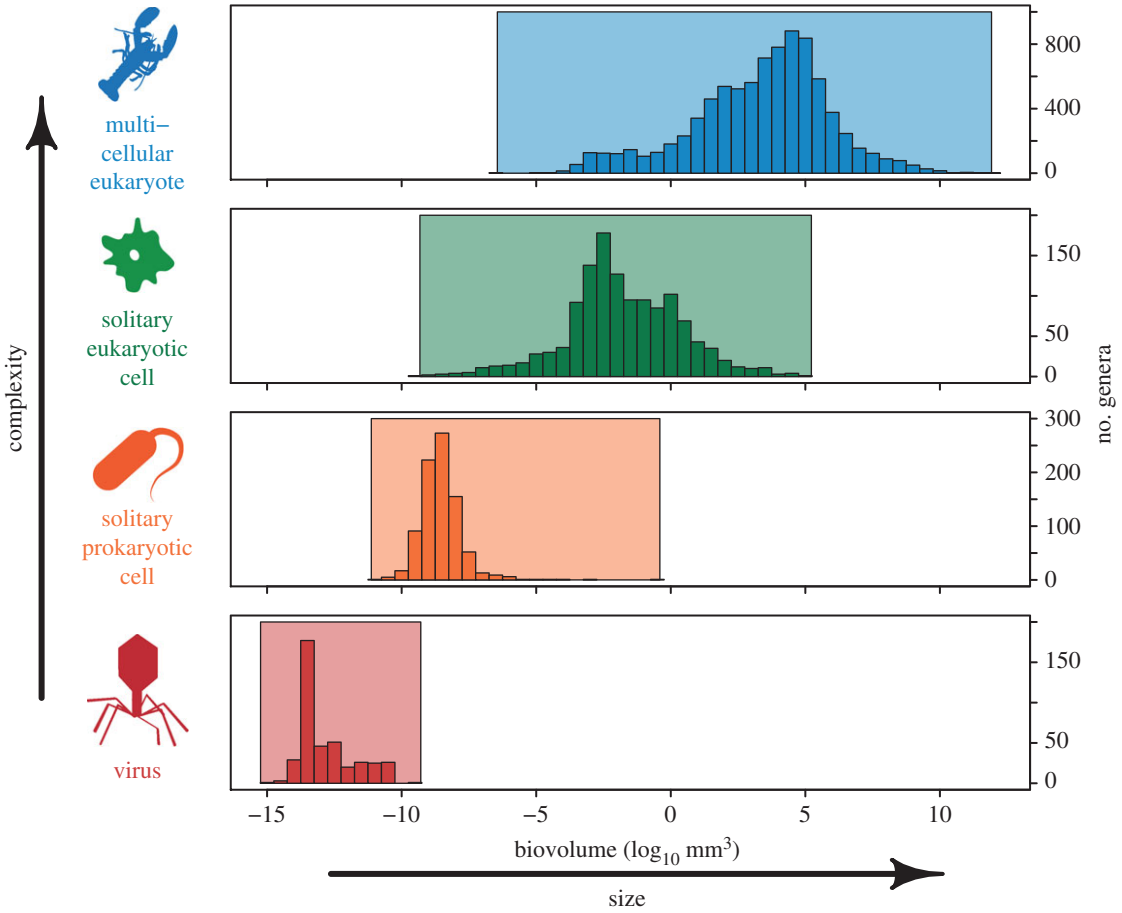

Figure 2. Distributions of extant organismal size within each of the vertical levels. Coloured shaded areas highlight the total size range occupied by living genera within a given level. As the hierarchy is ascended from viruses to multicellular eukaryotes, the modal size as well as the total range in size increases. (Online version in colour.)

Table 1. Descriptive statistics of the size distributions within each level of vertical complexity. All organism sizes (cubic millimetres) were $\log_{10}$ transformed before calculating statistics.

	n	min	max	range	midpoint	mean	median	s.d.	coefficient of variation
virus	405	−15.5	−9.5	6.0	−12.5	−13.1	−13.5	1.06	0.081
prokaryote	851	−11.4	−0.7	10.7	−6.1	−8.8	−8.8	0.80	0.091
protist	1284	−9.6	5.3	14.9	−2.2	−2.0	−2.2	2.14	1.070
multicellular eukaryote	8632	−6.7	11.7	18.4	2.5	3.1	3.5	2.55	0.816

relative diversity in each group, is statistically indistinguishable from unimodality (figure 1*d*), but this must be the case simply because of the extreme weighting of a single complexity level: prokaryotes. Neither the raw nor corrected size distributions reproduce the pattern predicted by Gould [12].

Three principal patterns emerge when comparing the size distributions of genera among the complexity levels. First, the mean biovolume within each level of complexity is five to eight orders of magnitude larger than at the next lower level; this is also true for medians and size-range midpoints (figure 2 and table 1; electronic supplementary material, figure S1). The tails of adjacent levels overlap—such that the largest 'simpler' organisms are larger than the smallest organisms of each subsequent level—but the modes are well separated (figure 2).

Second, the range of sizes within each level increases as complexity increases from viruses to multicellular eukaryotes (figure 2 and table 1; electronic supplementary material, figure S1). Ascending the complexity hierarchy, the total span of sizes increases by many orders of magnitude compared with the preceding level (table 1). Viruses span six orders of magnitude, whereas prokaryotes span nearly 11, protists span 19 and multicellular eukaryotes span 23. Moreover, the variances also differ significantly among levels (Brown–Forsythe test of equal variances; $F = 270.7$, $p \ll 0.001$); this result holds when comparing all levels simultaneously as well as for comparisons between adjacent vertical levels, such as prokaryotes and protists (table 2). The sole exception to this pattern is the comparison between viruses and prokaryotes, where viruses have a slightly larger variance. Data presented

Table 2. Results from Brown–Forsythe test of equal variances. The top row shows results for comparing all levels simultaneously. The bottom three rows show the results of comparisons between adjacent vertical levels.

	difference in variance	test statistic	*p*-value
all levels	n.a.	273.3	$\ll 0.0001$
virus versus prokaryote	−0.51	49.7	$\ll 0.0001$
prokaryote versus protist	3.97	497.1	$\ll 0.0001$
protist versus multicellular eukaryote	1.90	39.9	$\ll 0.0001$

here are $\log_{10}$ transformed to reduce statistical problems associated with differences in scale. However, the coefficient of variation (c_v) can be used to further correct for differences in mean values associated with each level (table 1). These results are similar to the raw sample standard deviations. The c_v increases from viruses to prokaryotes, but multicellular eukaryotes have a slightly smaller c_v than protists.

Third, there is a striking asymmetry between the increases in minimum and maximum size bounds (table 1). Minimum size increases by two-to-four orders of magnitude as each level of the hierarchy is breached, whereas the maximum size increases by six-to-nine orders of magnitude. For example, the minimum sizes of prokaryotes and protists differ by 1.8 $\log_{10}$ units while the maximum sizes of these two groups differ by 6.8 $\log_{10}$ units.

4. Discussion

The results indicate that the distribution of organismal sizes across the tree of life is multimodal. Moreover, the aggregate distribution is composed of distinct subgroups distinguished by levels of structural complexity: each level of complexity has its own unimodal size distribution, differing in minimum, median, mean and maximum size as well as size range from all other levels (figure 1*b*). Our analyses suggest that biological structure limits the range of potential sizes, consistent with previous suggestions that each level of complexity is bounded by minimum and maximum constraints [13].

For three out of four levels of complexity (virus, prokaryote and protist), we have compiled data from taxonomic compendia to ensure representative sampling of the intra-level size distributions. However, for the multicellular eukaryotes we relied on taxon-specific literature and database sources. Despite some unavoidable bias in data compilation, we have sampled across the size range of non-colonial multicellular eukaryotes from nematodes to mammals and trees. The most obvious bias in the multicellular eukaryote dataset is that, relative to the phylum with the largest number of genera in our dataset, the Mollusca, we have over-sampled Nematoda and under-sampled Arthropoda [24] because of the availability of appropriate taxonomic compendia (i.e. the 'monographic effect'). More proportionate sampling, however, is unlikely to qualitatively change the observed patterns. Increasing the representation of arthropods, particularly of insects, would certainly shift the mode of the size distribution towards smaller sizes, but it would still not overlap with the protist mode. When we examine a sample of fossil insect sizes alongside the extant multicellular and unicellular eukaryotes in our dataset, the modal size class of fossil insects falls between the modes for extant multicellular eukaryotes and protists observed in our data (electronic supplementary material, figure S2). Including the hyper-diverse Insecta is only likely to decrease the multicellular eukaryote mode by at most one or two $\log_{10}$ biovolume units; that is, the mode would remain distinct from the protist size mode.

The unimodality of sizes within each level (figure 2) suggests that organisms at each level of complexity must be substantially larger than those of the preceding level, largely because individuals are composed of multiple components from the lower level. This constraint does not apply to horizontal complexity, however, which allows continuous relationships between cell types and body size. It is also compatible with the observation that Linnaean classes of multicellular eukaryotes are symmetrically distributed individually and sum to a unimodal, approximately symmetric distribution (electronic supplementary material, figure S3). Our results suggest that there are constraints on both the minimum and maximum sizes for each level of complexity. We sampled the extreme sizes for all levels to ensure accurate representation of the overall known organismal size range (figure 2; electronic supplementary material, figure S1), but we emphasize that the current limits are not necessarily equivalent to the theoretical ones.

We can use the size distributions of fossil organisms to test the hypothesis that increasing complexity is necessary for large and rapid increases in size. The earliest genera preserved in the fossil record after each major complexity transition have approximately the same median and mean sizes as their modern counterparts and are much larger than their contemporaries at the next lower level (figure 3). The observed patterns may be distorted by differential fossilization of organisms in different size classes. For example, the fact that all pre-Statherian (greater than 1800 Myr) fossil prokaryotes fall within the large end of the modern cell-size distribution (figure 3) suggests a size bias in preservation. This apparent bias strengthens the argument that the size jump from prokaryotes to unicellular eukaryotes during the Statherian Period is much larger than the subsequent increase in median size of protists. Similarly, the increase in median size from unicellular eukaryotes to multicellular eukaryotes is larger than the subsequent increase in the median size of multicellular eukaryotes. Notably, the size distributions of Statherian (1800–1600 Myr), Cryogenian (850–635 Myr), Ediacaran (635–541 Myr) and modern protists are quite similar (figure 3). The similarity is particularly remarkable given that our samples of living taxa include many clades that lack mineralized tests and are typically not represented as fossils. Likewise, the median size of early animals during the Cambrian (541–485 Myr) is 0.8 $\log_{10}$ biovolume units smaller than the median size of modern multicellular eukaryotes. This increase in median size from the Cambrian to the recent is small compared with the 5.7 $\log_{10}$ biovolume unit difference between modern protists and multicellular eukaryotes. The fossil record of Ediacaran multicellular eukaryotes is dominated by large-bodied, stem-group animals whose relationships to crown-group phyla are uncertain [28], and

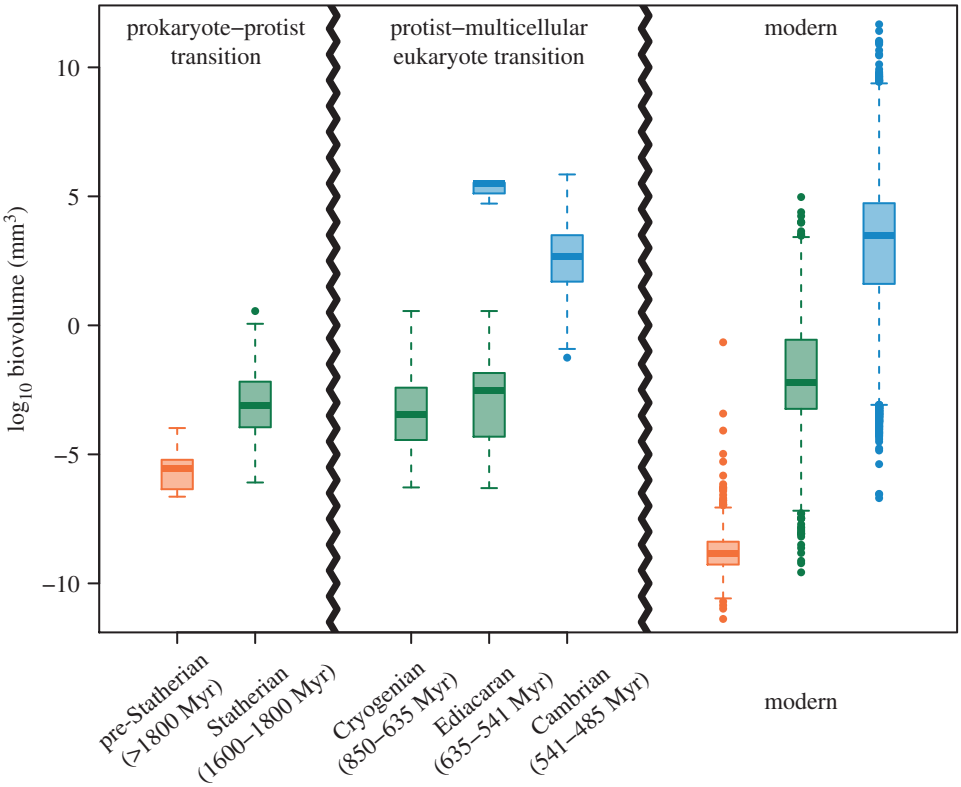

Figure 3. Box-and-whisker plots of organismal size distributions during major evolutionary transitions and the modern biota. (*a*) The transition from prokaryotes to eukaryotes during the Palaeoproterozoic era. Protists first appear in the fossil record during the Statherian period of the Palaeoproterozoic. The pre-Statherian prokaryotes are shown in orange for comparison. (*b*) Box-and-whisker plots for protists in the Cryogenic period, protists and multicellular eukaryotes during the Ediacarian period, and multicellular eukaryotes during the Cambrian. The first multicellular eukaryotes appear in the Ediacaran period and are represented mainly by large stem-group animals that are difficult to classify. Diverse representatives of crown-group animal phyla do not appear until the Cambrian period. (*c*) Box-and-whisker plots for the sizes of extant prokaryotes, protists and multicellular eukaryotes. Note that the first fossilized representatives of each complexity level are near the median size of their living counterparts and much larger than their contemporaries at the next lower level. Colour scheme follows figure 1: orange, prokaryote; green, protist; blue, multi-cellular eukaryote.

whose preservation occurred through a mode that is not common in Earth's history either before or after [29]. Together these factors likely explain the large median size of Ediacaran multicellular eukaryotes.

Our results are most consistent with the hypothesis that minimum size constraints arise from physical factors while constraints on maximum size arise from physiological factors [30,31]. This is almost certainly the case for prokaryotes, where the smallest species have just enough cell volume to contain the necessary genome, ribosomes and proteins to function as prokaryotes, whereas the largest species are limited by diffusion of materials across their surface area [15]. The largest known prokaryote, *Thiomargarita namibiensis* Schulz and others 1999, is approximately eight orders of magnitude larger in biovolume than the modal prokaryote. These extremely large prokaryotes have large vacuoles that occupy up to 98% of the cell's volume [32], thus limiting the metabolically active portion of the cell to a thin outer shell and shortening the distance over which materials must diffuse into and out of the cell. The extreme sizes of protists are likely to be limited by the same factors: a small cell must be large enough to contain all its necessary parts, while at large size transporting materials into, out of and within a cell becomes limiting. Protists, like all eukaryotes, use cytoplasmic streaming to facilitate intracellular transport. Animals appear to follow a similar pattern whereby the smallest animals, the parasitic myxozoans, are composed of only a few cells and possess highly reduced genomes [33]. The sizes of some of the largest multicellular eukaryotes, the baleen whales, are potentially constrained by a number of physiological factors such as thermoregulation [34] and the ability to acquire enough food [35,36].

Simple scaling calculations illustrate the metabolic challenges of being very large. Consider time scales of O_2 transport within a cell by both diffusion and cytoplasmic streaming, where the former scales with the square of length while the latter scales with length ([37] and references therein). It would take O_2 approximately 10^4 s to diffuse through a spherical prokaryote that is 1 mm in diameter. In a same-sized protist, mixing via cytoplasmic streaming at a typical streaming rate would take approximately 10^2 s. Scaling up the sizes of these hypothetical organisms two orders of magnitude to a diameter of 10 cm, the diffusion and mixing times would be approximately 10^7 and 10^4 s, respectively. These times for intracellular O_2 transport do not preclude the evolution of centimetre-scale bacteria or metre-scale protists, but such organisms would necessarily have extremely low metabolic rates. Intra-organism transport calculations are not so simple for multicellular organisms, which have evolved respiratory organs and circulatory systems. Nevertheless, there are metabolic constraints to being very large. For example, the energetic cost of lunge feeding increases with size such that whales larger than the largest known blue whale would require too much recovery time after each feeding lunge to meet the total metabolic demand [36].

Constraints on 'body' size for viruses may differ from those of cellular organisms because as non-cellular obligate parasites they do not have a self-sustained metabolism. Nonetheless, we

hypothesize that the constraint on minimum size in viruses is the same as that for the other organisms discussed here: they must be large enough to contain the minimum number of proteins to form a capsid and a minimum amount of genetic material to insert into a host's genome for replication. However, the constraint on maximum virus size is unlikely to be physiological, because viruses are not metabolically active and rely on their host for reproduction. Rather, the largest viruses appear to be limited by the sizes of their hosts [38]. Viruses are reproduced inside the cell of their host and so must be smaller than their host cell and possess a genome that is small relative to that of the host [38]. The recently discovered giant Kloneuviruses—which probably evolved from small viral ancestors by acquiring a large number of genes from cellular hosts [39]—provide another example of increasing size due to increasing horizontal complexity, though as obligate parasites they still must be small relative to their host cells.

Each hierarchical level exhibits a unimodal size distribution with its mean and median located near the midpoint of the range (figure 2; electronic supplementary material, figure S1). The generation of these size–frequency distributions most likely involves some form of taxon sorting [40], and, specifically, effect-macroevolution [41,42]. Effect-macroevolution is postulated to occur when the trait undergoing evolution is a property of the organism rather than an emergent property of a higher-level taxon, and the trait is associated with differential macroevolutionary dynamics at those higher levels (e.g. extinction rate covaries with body size). Maurer *et al.* [43] modelled effect-macroevolution in isolation and in combination with anagenetic evolution to explain the skewness observed in mammalian body size distributions across continents. They found that a variety of parameter combinations, including size-related biases in origination and extinction rates, can produce skewed and symmetric size–frequency distributions. More concretely, if organisms near the lower size limit of each vertical level tend to have reduced genomes, as do the parasitic myxozoans [33], then they may have less genetic variation, and hence lower speciation rates. Very large organisms also may have lower speciation rates because larger organisms typically have large geographical ranges [44], size-specific behavioural adaptations [45] or because they operate very close to their metabolic limits, as is the case with blue whales [36]. Importantly, in this view, these rate differentials are macroevolutionary properties, and there is no implication that individuals in species far from the mode are less fit than those nearer to it.

The effect-macroevolution hypothesis is difficult to test without comprehensive phylogenies or robust fossil data for all levels of complexity, but it is consistent with the explanation for size evolution in terrestrial mammals over the Cenozoic [46] and marine animals over the Phanerozoic [47]. Given the taxonomic scope and the relatively low proportion of total diversity sampled in the present study, quantitative assessment of skewness is beyond the scope of our available data. Nevertheless, taxon sorting by effect-macroevolution is one potential explanation for the shapes of the size–frequency distributions across hierarchical levels [43].

Alternatively, the size distributions observed here could result from purely stochastic variations in origination and extinction rates with no covariation between size and macroevolutionary rates, in other words, simply from increasing variance. Such a process was hypothesized by Stanley [10] (see also [11,12]), who proposed that size evolved as an increase in variance without any size-related biases in macroevolutionary processes. The resulting distribution, if free of or distant from constraints, would be unimodal, symmetric and centred on the ancestral size. There is some support in our data for the increasing-variance hypothesis within vertical levels. At all three complexity levels for which we have fossil data, the sizes of the first species in the fossil record are very close to the modal size of their modern counterparts (figure 3). If the increasing variance model holds, then the current size distributions are evolutionarily contingent: the distributions resulted from a diffusion-like evolutionary process initialized at the ancestral size. Diffusion-like processes operating within each hierarchical level do not imply that there is no selection. At the higher level of vertical complexity, the multitude of selective processes operating at lower levels may produce the emergent patterns that resemble diffusion [10].

Regardless of whether or not diffusive evolution is the primary mode operating within complexity levels, we can test for differences and commonalities in process among the levels. Under a model of diffusive evolution, the range of sizes observed among crown groups is determined by the time origination and the rate of speciation [48]. If we assume size changes between ancestors and descendants are sampled from a single distribution with a time-invariant standard deviation and further assume constant speciation rates across all clades, we would expect older clades to occupy larger size ranges. Similarly, if we assume variable rates of speciation but equal evolutionary age, we would expect clades with faster rates to occupy larger size ranges.

Our data on the total size ranges occupied at each level of complexity are not consistent with either of these predictions. The two oldest groups considered here are the viruses and prokaryotes. The age of the first virus is not known because viruses lack a fossil record, but given that they are obligate parasites on cellular organisms, the simplest assumption is that they postdate the prokaryotes, which originated approximately 3.8 Ga. The fossil record shows that solitary eukaryotes first appeared approximately 1.9 Ga, followed by metazoans at or shortly before 0.6 Ga. Given this evolutionary sequence, we would expect prokaryotes and viruses to occupy the largest range in size, followed by protists and then multicellular eukaryotes (figure 2; electronic supplementary material, figure S1). This prediction is not borne out by the data. The smallest range in size is occupied by the viruses, followed by prokaryotes, protists and multicellular eukaryotes.

Considering molecular evolutionary rates, again we find discordance between the expected and observed outcomes. Viruses have the highest per nucleotide per generation mutation rates [49] and the smallest range in size (figure 2 and table 1; electronic supplementary material, figure S1). At the other end of the hierarchy, multicellular eukaryotes have the largest size range (figure 2 and table 1; electronic supplementary material, figure S1), but mutation rates intermediate between those of prokaryotes and those of viruses [49]. The size distribution data presented here (figures 1, 2 and table 1; electronic supplementary material, figure S1) do not support uniform rates of size evolution across the complexity hierarchy.

We find that viruses conform to the overall patterns exhibited by cellular life despite ambiguity over whether they are living organisms. Viruses are unusual in that they probably arose by a decrease in vertical complexity from cellular ancestors [50,51]. Partial decreases in vertical

complexity—what have been called the 'minor transitions' in evolution [52], such as the transitions from simple multicellularity to solitary living in certain algae—are not uncommon. In fact, decreases of this sort were slightly more common than partial increases [52]. Larger downward transitions—full drops from one vertical level to the next one down—appear to have occurred multiple times. The transitions from bacteria to viruses are examples of this [51], as is the evolution of clonally transmissible cancers in dogs and Tasmanian devils [53]. Nevertheless, the size range of viruses conforms to the pattern of smaller size and reduced size range relative to those of the next higher level (figure 2; electronic supplementary material, figure S1). This observation suggests that size ranges within levels of vertical organization are bounded by real biological limits, whether physical or physiological, that are independent of phylogeny. Thus, a derived state of reduced vertical complexity does not carry the ancestral size constraints of the previously occupied vertical level. In other words, the size of a virus that evolved from a cellular organism is constrained by the complexity associated with being a virus and not by the shared evolutionary history with much larger cellular organisms.

Of course, there are potential constraints on organismal size other than vertical complexity. The most obvious is predation pressure, which has been argued to drive size increases in both prey and predator species. For example, escalatory 'arms races' have been implicated in driving increases in body size and the diversification of morphological defence mechanisms during the Mesozoic marine revolution [54]. Such dynamics have most likely been important in shaping the size distributions of particular clades within levels of hierarchical complexity (e.g. families of gastropods), but they do not explain why each level exhibits an overall unimodal size distribution despite complex interactions among component taxa. Because extant life spans 23 orders of magnitude, species that are too small or too large to be consumed by other species are exceedingly rare.

The size distributions from the fossil record demonstrate large jumps in size when new levels of complexity evolve, followed by increase in variance around the initial median size (figure 3). These jumps are not simple shifts in central tendency (mean, median and mode), but rather fundamental changes in the shape of size distributions, including their dispersions, minima and maxima. This observation provides strong evidence that vertical complexity has been the major factor limiting the sizes of organisms across all grades of life. The earliest members of each new complexity level quickly reach the mode for that level, and the total size range within a level increases over evolutionary time as new species evolve mechanisms for exploiting body size niches far from the modal size.

In summary, body size data spanning all major branches on the tree of life reveal discrete macroevolutionary modes for different levels of vertical complexity. The individual size distributions for multicellular eukaryotes, protists, prokaryotes and viruses are approximately unimodal and symmetric, but the combined distribution is multimodal and highly asymmetric. Evolutionary innovations associated with new levels of complexity therefore appear to be fundamentally different in nature from those that arise within complexity levels.

Ethics. Data for this study were derived from the published literature. No living or dead animal (inclusive of *Homo sapiens*), plant, protist, bacteria or virus subjects/specimens were used in this research.

Data accessibility. All data used in analyses and figures are included as the electronic supplementary material and are also freely accessible at the Stanford Digital Repository (https://purl.stanford.edu/db950qy6691).

Authors' contributions. This study was designed by N.A.H. and J.L.P. Analyses were performed by N.A.H. and he wrote the initial draft of the manuscript, and all authors contributed to subsequent revisions. All authors made significant intellectual contributions to the final manuscript.

Competing interests. We declare we have no competing interests.

Funding. This study is a product of the working group on body size evolution (principal investigators J. Payne, J. Stempien and M. Kowalewski) funded by the National Evolutionary Synthesis Center (NESCent), National Science Foundation grant EF-0423641. Additional funding was provided by a CAREER grant from the US National Science Foundation to J.L.P. (EAR-1151022).

References

1. Kowalewski M *et al.* 2011 The Geozoic supereon. *Palaios* **26**, 251–255. (doi:10.2110/palo.2011.S03)
2. Payne JL *et al.* 2009 Two-phase increase in the maximum size of life over 3.5 billion years reflects biological innovation and environmental opportunity. *Proc. Natl Acad. Sci. USA* **106**, 24–27. (doi:10.1073/pnas.0806314106)
3. McShea DW. 2001 The hierarchical structure of organisms: a scale and documentation of a trend in the maximum. *Paleobiology* **27**, 405–423. (doi:10.1666/0094-8373(2001)027<0405:THSOOA>2.0.CO;2)
4. Sterelny K. 1999 Bacteria at the high table. *Biol. Philos.* **14**, 459–470. (doi:10.1023/A:1006542531480)
5. Sagan L. 1967 On the origin of mitosing cells. *J. Theor. Biol.* **14**, 225–274. (doi:10.1016/0022-5193(67)90079-3)
6. Richter DJ, King N. 2013 The genomic and cellular foundations of animal origins. *Annu. Rev. Genet.* **47**, 509–537. (doi:10.1146/annurev-genet-111212-133456)
7. Valentine JW. 2004 *On the origin of phyla*. Chicago, IL: University of Chicago Press.
8. McShea DW. 1996 Metazoan complexity and evolution: is there a trend? *Evolution* **50**, 477–492.
9. McShea DW. 1994 Mechanisms of large-scale evolutionary trends. *Evolution* **48**, 1747–1763. (doi:10.2307/2410505)
10. Stanley SM. 1973 An explanation for Cope's Rule. *Evolution* **27**, 1–26. (doi:10.1111/j.1558-5646.1973.tb05912.x)
11. Gould SJ. 1988 Trends as changes in variance: a new slant on progress and directionality in evolution. *J. Paleontol.* **62**, 319–329. (doi:10.1017/S0022336000059126)
12. Gould SJ. 1996 *Full house: the spread of excellence from Darwin to Plato*. New York, NY: Harmony Books.
13. Knoll AH, Bambach RK. 2000 Directionality in the history of life: diffusion from the left wall or repeated scaling of the right? *Paleobiology* **26**, 1–14. (doi:10.1666/0094-8373(2000)26[1:DITHOL]2.0.CO;2)
14. Knoll AH. 1999 *Size limits of very small microorganisms: proceedings of a workshop*. Washington, DC: National Academy Press.
15. Koch AL. 1996 What size should a bacterium be? A question of scale. *Annu. Rev. Microbiol.* **50**, 317–348. (doi:10.1146/annurev.micro.50.1.317)
16. Bell G, Mooers AØ. 1997 Size and complexity among multicellular organisms. *Biol. J. Linn. Soc. Lond.* **60**, 345–363. (doi:10.1111/j.1095-8312.1997.tb01500.x)
17. Bonner JT. 2004 The size–complexity rule. *Evolution* **58**, 1883–1890. (doi:10.1111/j.0014-3820.2004.tb00476.x)
18. Raup DM, Boyajian GE. 1988 Patterns of generic extinction in the fossil record. *Paleobiology* **14**, 109–125. (doi:10.1017/S0094837300011866)

19. Gastwirth JL, Gel YR, Hui WLW, Lyubchich V, Miao W, Noguchi K. 2015 lawstat: tools for biostatistics, public policy, and law. R package version 3.0. See https://CRAN.R-project.org/package=lawstat.
20. R Core Team. 2016 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
21. Hartigan JA, Hartigan PM. 1985 The dip test of unimodality. *Ann. Stat.* **13**, 70–84. (doi:10.2307/2241144)
22. Maechler M. 2015 diptest: Hartigan's dip test statistic for unimodality—corrected. R package version 0.75-7. See https://CRAN.R-project.org/package=diptest.
23. Roy K, Jablonski D, Valentine JW. 1996 Higher taxa in biodiversity studies: patterns from Eastern Pacific marine molluscs. *Phil. Trans. R. Soc. Lond. B* **351**, 1605–1613. (doi:10.1098/rstb.1996.0144)
24. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011 How many species are there on Earth and in the ocean? *PLoS Biol.* **9**, e1001127. (doi:10.1371/journal.pbio.1001127)
25. May RM. 1988 How many species are there on Earth? *Science* **241**, 1441–1449. (doi:10.1126/science.241.4872.1441)
26. Rohwer F. 2003 Global phage diversity. *Cell* **113**, 141. (doi:10.1016/S0092-8674(03)00276-9)
27. Kowalewski M, Finnegan S. 2010 Theoretical diversity of the marine biosphere. *Paleobiology* **36**, 1–15. (doi:10.1666/0094-8373-36.1.1)
28. Xiao S, Laflamme M. 2009 On the eve of animal radiation: phylogeny, ecology and evolution of the Ediacara biota. *Trends Ecol. Evol.* **24**, 31–40. (doi:10.1016/j.tree.2008.07.015)
29. Gehling JG. 1999 Microbial mats in terminal Proterozoic siliciclastics: Ediacaran death masks. *Palaios* **14**, 40–57. (doi:10.2307/3515360)
30. Smith FA, Lyons SK. 2013 On being the right size. In *Animal body size* (eds FA Smith, SK Lyons), pp. 1–10. Chicago, IL: University of Chicago Press.
31. Maurer BA, Marquet PA. 2013 Processes responsible for patterns in body mass distributions. In *Animal body size* (eds FA Smith, SK Lyons), pp. 168–186. Chicago, IL: University of Chicago Press.
32. Schulz HN. 2006 The genus *Thiomargarita*. In *The prokaryotes: protobacteria: gamma subclass*, vol. 6 (eds M Dworkin, S Falkow, E Rosenberg, K-H Scleifer, E Stackebrandt), pp. 1156–1163. New York, NY: Springer.
33. Chang ES *et al.* 2015 Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. *Proc. Natl Acad. Sci. USA* **112**, 14 912–14 917. (doi:10.1073/pnas.1511468112)
34. Hokkanen JEI. 1990 Temperature regulation of marine mammals. *J. Theor. Biol.* **145**, 465–485. (doi:10.1016/S0022-5193(05)80482-5)
35. Goldbogen JA, Potvin J, Shadwick RE. 2010 Skull and buccal cavity allometry increase mass-specific engulfment capacity in fin whales. *Proc. R. Soc. B* **277**, 861–868. (doi:10.1098/rspb.2009.1680)
36. Potvin J, Goldbogen JA, Shadwick RE. 2012 Metabolic expenditures of lunge feeding rorquals across scale: Implications for the evolution of filter feeding and the limits to maximum body size. *PLoS ONE* **7**, e44854. (doi:10.1371/journal.pone.0044854.g016)
37. Payne JL, Groves JR, Jost AB, Nguyen T, Moffitt SE, Hill TM, Skotheim JM. 2012 Late Paleozoic fusulinoidean gigantism driven by atmospheric hyperoxia. *Evolution* **66**, 2929–2939. (doi:10.1111/j.1558-5646.2012.01626.x)
38. Claverie J-M, Ogata H, Audic S, Abergel C, Suhre K, Fournier P-E. 2006 Mimivirus and the emerging concept of 'giant' virus. *Virus Res.* **117**, 133–144. (doi:10.1016/j.virusres.2006.01.008)
39. Schulz F *et al.* 2017 Giant viruses with an expanded complement of translation system components. *Science* **356**, 82–85. (doi:10.1126/science.aal4657)
40. Vrba ES, Gould SJ. 1986 The hierarchical expansion of sorting and selection: sorting and selection cannot be equated. *Paleobiology* **12**, 217–228. (doi:10.1017/S0094837300013671)
41. Vrba ES. 1983 Macroevolutionary trends: new perspectives on the roles of adaptation and incidental effect. *Science* **221**, 387–389. (doi:10.1126/science.221.4608.387)
42. Jablonski D. 2008 Species selection: theory and data. *Annu. Rev. Ecol. Evol. Syst.* **39**, 501–524. (doi:10.1146/annurev.ecolsys.39.110707.173510)
43. Maurer BA, Brown JH, Rusler RD. 1992 The micro and macro in body size evolution. *Evolution* **46**, 939–953. (doi:10.1111/j.1558-5646.1992.tb00611.x)
44. Martin RA. 2016 Body size in (mostly) mammals: mass, speciation rates and the translation of gamma to alpha diversity on evolutionary timescales. *Hist. Biol.* **29**, 576–593. (doi:10.1080/08912963.2016.1211646)
45. Liow LH, Fortelius M, Bingham E, Lintulaakso K, Mannila H, Flynn L, Stenseth NC. 2008 Higher origination and extinction rates in larger mammals. *Proc. Natl Acad. Sci. USA* **105**, 6097–6102. (doi:10.1073/pnas.0709763105)
46. Smith FA *et al.* 2004 Similarity of mammalian body size across the taxonomic hierarchy and across space and time. *Am. Nat.* **163**, 672–691. (doi:10.1086/382898)
47. Heim NA, Knope ML, Schaal EK, Wang SC, Payne JL. 2015 Cope's rule in the evolution of marine animals. *Science* **347**, 867–870. (doi:10.1126/science.1260065)
48. Trammer J. 2005 Maximum body size in a radiating clade as a function of time. *Evolution* **59**, 941–947. (doi:10.1111/j.0014-3820.2005.tb01033.x)
49. Lynch M. 2010 Evolution of the mutation rate. *Trends Genet.* **26**, 345–352. (doi:10.1016/j.tig.2010.05.003)
50. Forterre P, Krupovic M. 2012 The origin of virions and virocells: the escape hypothesis revisited. In *Viruses: essential agents of life* (ed. G Witzany), pp. 43–60. Dordrecht, The Netherlands: Springer.
51. Nasir A, Sun FJ, Kim KM, Caetano Anollés G. 2015 Untangling the origin of viruses and their impact on cellular evolution. *Ann. NY Acad. Sci.* **1341**, 61–74. (doi:10.1111/nyas.12735)
52. Marcot JD, McShea DW. 2007 Increasing vertical complexity throughout the history of life: phylogenetic tests of trend mechanisms. *Paleobiology* **33**, 182–200. (doi:10.1666/06028.1)
53. Murchison EP. 2009 Clonally transmissible cancers in dogs and Tasmanian devils. *Oncogene* **27**, S19–S30. (doi:10.1038/onc.2009.350)
54. Vermeij GJ. 1977 The Mesozoic marine revolution: evidence from snails, predators and grazers. *Paleobiology* **3**, 245–258. (doi:10.1017/S0094837300005352)